Bayesian inference of multivariate-GARCH-BEKK models

G. C. Livingston Jr[1] · Darfiana Nur[2]

Abstract

The main aim of this paper is to present a Bayesian analysis of Multivariate GARCH(l, m) (M-GARCH) models including estimation of the coefficient parameters as well as the model order, by combining a set of existing MCMC algorithms in the literature. The proposed algorithm focuses on the BEKK formulation of the multivariate GARCH model. The estimation procedure will be designed as a custom MCMC with embedded Reversible Jump MCMC (RJMCMC) and Delayed Rejection Metropolis-Hastings (DRMH) steps implemented using the statistical software R. The RJMCMC steps allow three variants of BEKK models (constant, diagonal and full) to be indexed and this index included as a parameter to be estimated. The proposed MCMC algorithms are validated using extensive simulation experiments followed by a case study using bivariate data derived from the daily share prices for BHP Group Limited, Rio Tinto Group, and Fortescue Metals Group Limited on the ASX over from September 2013 to December 2021.

Keywords Gibbs Sampler · Metropolis-Hastings · Reversible Jump MCMC · Delayed Rejection MH · MGARCH

1 Introduction

The model that is most commonly suggested in the literature to capture changing volatility in time series is the Generalised Autoregressive Conditional Heteroskedastic (GARCH) model (Bollerslev 1986). For univariate GARCH models, there are several representations of the conditional variance equation. Some examples are the regu-

✉ G. C. Livingston Jr
  glen.livingstonjr@newcastle.edu.au

1 Centre for Computer-Assisted Research Mathematics and Its Applications, School of
   Information and Physical Sciences, The University of Newcastle, University Drive, Callaghan,
   NSW 2308, Australia

2 School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Kent
   Street, Bentley, WA 6102, Australia

lar GARCH (Bollerslev 1986), the EGARCH (Nelson 1991) and the GJR-GARCH (Glosten et al. 1993) models.

Within a multivariate setting it is just as important, if not more so, to allow the conditional covariance to change over time. By doing so, it is possible to gauge how changes in volatility for one series affect the volatility of another. This transmission of volatility from one series to another is sometimes referred to as a spill over effect. Financial contagion is an example of this type of effect. For the multivariate GARCH (M-GARCH) model, as one might expect, there are a number of different representations of the conditional covariance equation. There have been several valuable reviews and surveys of the rapidly changing literature on M-GARCH models. Some examples are Bauwens et al. (2006), Bauwens et al. (2012), Silvennoinen and Teräsvirta (2009), Francq and Zakoïan (2010) and Tsay (2013).

Under a multivariate setting, the number of parameters requiring estimation is significantly increased. Assuming that both model orders are equal to one, the number of parameters in the bivariate GARCH model under the VECH formulation (Bollerslev et al. 1988) is 21. When three series are analysed, the number of parameters that must be estimated is increased to 78. This is a considerable increase over the number of parameters for two or three separate univariate GARCH models, and is an example of the curse of dimensionality. As a result of the increased number of parameters, the amount of data required to perform meaningful estimation increases, which, in turn, makes estimation by numerical methods more time consuming and difficult.

Another formulation of the conditional covariance equation was proposed by the Engle and Kroner (1995) and is referred to as the BEKK model. The name is an acronym made out of the initials of those who contributed to the development of M-GARCH models, namely, Baba, Engle, Kraft and Kroner. The BEKK formulation of the conditional covariance equation ensures that the conditional covariance equations are positive definite and also has less parameters than the VECH formulation.

Practical applications of the BEKK formulation lie in the fields of finance and economics. Often, the subject of investigation is the transfer of volatility between markets. Kearney and Patton (2000) looked at the transfer of volatility in exchange rate data. A spill over of volatility from the Hong Kong stock exchange to the exchanges in mainland China was found to be unidirectional in a study by Li (2007) that applied the BEKK model. Saleem (2009) measured the effect of the 1998 Russian financial crisis on other international financial markets using several bivariate BEKK models.

There has been a comparatively small amount of literature produced on M-GARCH models through a Bayesian setting. Several multivariate ARCH and GARCH type models were applied to bivariate foreign exchange rate data in Vrontos et al. (2003). Differences in inference between the Bayesian method and classical estimation were found for scenarios in which the posterior distribution was non-normal. The posterior predictive distribution was used to compare the applied models: an M-GARCH model with diagonal covariance was found to be preferable, based on one step ahead prediction. A Bayesian non-parametric modelling approach for M-GARCH models was proposed by Jensen and Maheu (2013). Their approach was applied to a ten asset portfolio, with model selection performed using Bayes factors. They found that their best semi-parametric model performed significantly better in density forecasts than

models with Student-*t* innovations. The paper also provides guidance on expanding their approach to include a VAR model as the conditional mean.

The main aim of this paper is to present a Bayesian analysis of M-GARCH models which includes parameter estimation and model section into the one task. The algorithm combines a set of existing MCMC algorithms in the literature. The estimation procedure will be designed as a custom MCMC with embedded Reversible Jump MCMC (RJMCMC) (Green (1995)) and Delayed Rejection Metropolis-Hastings (DRMH) (Tierney and Mira (1999); Metropolis et al. (1953); Hastings (1970)) steps implemented using the statistical software R (R Development Core Team (2012)). The RJMCMC step allows different variants of the BEKK model to be included as a parameter to be estimated as part of the procedure. This removes the need for estimation of multiple models with differing conditional covariance equations, followed by a comparison and selection based on a scoring technique such as Deviance Information Criterion (DIC) (Spiegelhalter et al. (2002)). The DRMH steps are used for the coefficient parameter vectors, implemented in two stages which involves the prescription of proposal distributions in both stages of the algorithm for a faster convergence.

To achieve the aim of this paper, the organization is outlined as follows. Section 2 presents the methods including the details of the model, the prior distributions, the joint likelihood function and the joint posterior distribution function. Following this, the proposed MCMC algorithms and the design of simulation experiments will be detailed in Sect. 3. The results including posterior summaries of the parameters for extensive simulation studies and a case study are presented in Sect. 4. Finally, summary and future work are discussed in Sect. 5.

2 Methods

2.1 The M-GARCH model

The structure of M-GARCH models is similar to that of the univariate GARCH models presented in Bollerslev (1986). For a given vector of data x_t, the structure outlined in Bauwens et al. (2006) for a model with zero conditional mean is as follows:

$$
\begin{aligned}
x_t &= \varepsilon_t \\
\varepsilon_t &= H_t^{\frac{1}{2}} \eta_t.
\end{aligned}
\tag{1}
$$

Here, η_t is an independent identically distributed random vector of length p such that $E[\eta] = 0$ and $E[\eta \eta^T] = \mathbf{I}_p$. The conditional covariance matrix at time t is given by $H_t \in \mathbb{R}^{p \times p}$, and is required to be positive definite for all t. The $p \times 1$ vector ε_t is a vector of errors for the process. Under the assumption that η is distributed as a multivariate normal distribution, the errors of the process are distributed via a multivariate normal distribution, that is, $\varepsilon_t \sim \mathcal{N}_p(\mathbf{0}, H_t)$.

To derive the likelihood function and posterior distributions, it is best to write the model in matrix form as follows:

$$\mathbf{X} = \mathbf{E}$$
$$\text{vec}\left(\mathbf{E}^T\right) = \mathbf{H}^{\frac{1}{2}}\boldsymbol{\eta},$$

where

$$
\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_{s+1}^T \\ \boldsymbol{x}_{s+2}^T \\ \vdots \\ \boldsymbol{x}_N^T \end{bmatrix}, \quad
\mathbf{E} = \begin{bmatrix} \boldsymbol{\varepsilon}_{s+1}^T \\ \boldsymbol{\varepsilon}_{s+2}^T \\ \vdots \\ \boldsymbol{\varepsilon}_N^T \end{bmatrix}, \quad
\mathbf{H} = \begin{bmatrix} H_s & 0 & \cdots & 0 \\ 0 & H_{s+1} & \cdots & 0 \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & H_N \end{bmatrix}. \tag{2}
$$

The matrix $\mathbf{H} \in \mathbb{R}^{np \times np}$ in (2) is a block diagonal matrix in which each H_i is calculated using the conditional covariance equation relevant to the M-GARCH model being employed for $i = s, \ldots, N$. The other matrices in (2) have the following dimensions: $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{E} \in \mathbb{R}^{n \times p}$. In the following section, the BEKK formulation is described.

2.2 BEKK conditional covariance equation

The conditional covariance equation for the full BEKK model was originally defined as follows:

$$
H_t = C_0 C_0^T + \sum_{q=1}^{Q} \sum_{i=1}^{l} A_{qi} \boldsymbol{\varepsilon}_{t-i} \boldsymbol{\varepsilon}_{t-i}^T A_{qi}^T + \sum_{q=1}^{Q} \sum_{j=1}^{m} B_{qj} H_{t-j} B_{qj}^T, \tag{3}
$$

where A_{qi}, B_{qj} and C_0 are $p \times p$ parameter matrices, with C_0 being lower triangular. The parameters l and m are the model orders for the ARCH and GARCH parts of the model. The parameter Q is included in the model to accommodate more general representations of the conditional covariance equation. When $Q > 1$, an identification problem arises due to the existence of several parametrisations that lead to the same model. Conditions to eliminate redundant, observationally equivalent, models were provided by Engle and Kroner (1995).

Silvennoinen and Teräsvirta (2009) suggested that, due to numerical difficulties in estimation of BEKK models, it is often assumed that $l = m = Q = 1$. In fact in many applications of BEKK models, the simpler diagonal BEKK model is employed. Based on the advice of Silvennoinen and Teräsvirta (2009), we also restrict our model orders to those satisfying $l = m = Q = 1$. In addition, (3) shall be redefined so that the intercept matrix is defined directly to be a positive definite matrix, $C \in \mathbb{R}^{p \times p}$. In repeated trials, we found that it was very difficult to estimate the intercept matrix C_0 due to an identifiability issue caused by the calculation of $C_0 C_0^T$. We shall, consequently, employ the following form of the BEKK conditional covariance equation in this paper:

$$
H_t = C + A \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^T A^T + B H_{t-1} B^T. \tag{4}
$$

Table 1 Index of M-GARCH models

Index	$\mathcal{N}_1$	$\mathcal{N}_2$	$\mathcal{N}_3$
Model	Constant: $l = 0, m = 0$	DBEKK: $l = 1, m = 1$	BEKK: $l = 1, m = 1$

The number of parameters requiring estimation for the full BEKK model shown in (4) is $(l + m) \, p^2 Q + \frac{p}{2} (p + 1)$. While this is fewer than for a VECH model, even for small p, it is still a considerable number. The diagonal BEKK (DBEKK) model is a simpler form of the full BEKK model in which the coefficient parameter matrices A and B are diagonal. This form of the BEKK model has $(l + m) \, p Q + \frac{p}{2} (p + 1)$ parameters that require estimation. For some sets of data, for example those in which there is little transmission of volatility between series, the DBEKK model will be sufficient to account for the conditional covariance.

Rather than searching over BEKK models of differing model orders l and m, we shall search over a different variants of the conditional covariance equation. They are, a constant conditional covariance model, a DBEKK model and a full BEKK model, as shown in Table 1. The purely constant conditional covariance model is not a particularly interesting model to fit to time series data. However, it has been included here as it may be interesting if a model for the conditional mean was to be introduced. Use of a model index to search over is in line with the method used in Vrontos et al. (2000) and Livingston (2018).

The parameters that require estimation for the DBEKK and BEKK models are the intercept matrix C and the coefficient matrices A and B. For the constant conditional covariance model, only the intercept matrix C is estimated. The prior distributions are outlined in the following section.

2.3 Prior distribution

For the purposes of estimation, the coefficient matrices will be combined into a single vector of matrices, $\mathbf{A} = [A, B]^T$. We shall decompose the prior distribution for the M-GARCH estimation algorithm as follows:

$$\mathrm{p} \left(\mathcal{N}_j, \mathbf{A}, C \right) = \mathrm{p} \left(\mathbf{A}, C | \mathcal{N}_j \right) \mathrm{p} \left(\mathcal{N}_j \right).$$

The individual prior distributions are

$$\mathrm{p} \left(\mathbf{A}, C | \mathcal{N}_j \right) \propto \mathbb{I}_{\mathcal{A}} \left(\mathbf{A} \right)$$
$$\mathrm{p} \left(\mathcal{N}_j \right) \propto \frac{1}{j^{\tau_{\mathcal{N}}}}, \tag{5}$$

where j is the model index.

Stationarity will be enforced through the prior distribution of the coefficient matrices. This is achieved using the indicator function, $\mathbb{I}_{\mathcal{A}} \left(\mathbf{A} \right)$. This function takes the value of one when the conditions for covariance stationarity are met, and zero otherwise.

The BEKK model will be covariance stationary when the eigenvalues of the matrix

$$\sum_{q=1}^{Q}\sum_{i=1}^{l} A_{qi} \otimes A_{qi} + \sum_{q=1}^{Q}\sum_{j=1}^{m} B_{qj} \otimes B_{qj},$$

all have modulus less than one (Engle and Kroner 1995).

The prior distribution for the model type is defined in a way that allows simpler models with lower indices (j) to have greater prior probabilities. This obviously requires the models to be indexed from the simplest model to most complex. Setting the parameter $\tau_{\mathcal{N}} = 0$ incorporates the case where all prior model probabilities are proportional to one into the prior. In practice, the prior distribution can be set to any vector of prior probabilities that the user decides upon.

Now that we have considered the prior distributions, the next step is to derive the likelihood function for the M-GARCH model.

2.4 Likelihood function

The derivation of the likelihood function hinges upon the distribution of η_t. When η_t is distributed from a multivariate normal distribution, the likelihood function is the product of n multivariate normal densities with zero mean vector and covariance matrix H_t.

In general, if l and m are model orders for the conditional covariance equation, the likelihood function will be calculated using the first $s = \max(l, m)$ data points as initial values. The data available for the likelihood calculation is therefore indexed from $s + 1$ to N and the useful data has length $n = N - s$. The conditional likelihood is calculated as follows:

$$p\left(\mathbf{X}|\mathcal{N}_j, \mathbf{A}, C\right) = \prod_{t=(s+1)}^{N} p\left(x_t|\mathcal{N}_j, \mathbf{A}, C\right)$$

$$= (2\pi)^{-\frac{np}{2}} \left[\prod_{t=(s+1)}^{N} |H_t|\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{t=(s+1)}^{N} x_t^T H_t^{-1} x_t\right]. \tag{6}$$

Focusing on the sum in the exponent in (6), we calculate

$$\sum_{t=(s+1)}^{N} x_t^T H_t^{-1} x_t = x_{s+1}^T H_{s+1}^{-1} x_{s+1} + x_{s+2}^T H_{s+2}^{-1} x_{s+2} + \cdots + x_N^T H_N^{-1} x_N$$

$$= \left[x_{s+1}^T, x_{s+2}^T, \ldots, x_N^T\right] \left[\mathbf{H}^{-1}\right] \left[x_{s+1}, x_{s+2}, \ldots, x_N\right]^T \tag{7}$$

$$= \left[\text{vec}\left(\mathbf{X}^T\right)\right]^T \left[\mathbf{H}^{-1}\right] \left[\text{vec}\left(\mathbf{X}^T\right)\right].$$

The matrix $\mathbf{H}$ in (7) is a block diagonal matrix in which the ith diagonal entry corresponds to H_i. Substituting the expression for the sum from equation (7) back into the likelihood function gives us the final likelihood function for the M-GARCH model:

$$p\left(\mathbf{X}|\mathcal{N}_j, \mathbf{A}, C\right) = (2\pi)^{-\frac{np}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[\text{vec}\left(\mathbf{X}^T\right)\right]^T \left[\mathbf{H}^{-1}\right]\left[\text{vec}\left(\mathbf{X}^T\right)\right]\right]. \quad (8)$$

2.5 Posterior distribution

Combining the likelihood function in (8) with the prior distributions in (5) results in the following joint posterior distribution:

$$\begin{aligned}
p\left(\mathcal{N}_j, \mathbf{A}, C|\mathbf{X}\right) &\propto p\left(\mathbf{X}|\mathcal{N}_j, \mathbf{A}, C\right) p\left(\mathcal{N}_j, \mathbf{A}, C\right) \\
&\propto (2\pi)^{-\frac{np}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[\text{vec}\left(\mathbf{X}^T\right)\right]^T \left[\mathbf{H}^{-1}\right]\left[\text{vec}\left(\mathbf{X}^T\right)\right]\right] \frac{\mathbb{I}_{\mathcal{A}}\left(\mathbf{A}\right)}{j^{\tau_{\mathcal{N}}}}.
\end{aligned} \quad (9)$$

The parameters that are of interest are the model index, $\mathcal{N}_j$, the collection of coefficient parameter matrices $\mathbf{A}$ and the intercept matrix C. As we move from one model index to another, the dimension of the parameter space changes. Therefore, an RJMCMC algorithm is required for parameter estimation.

The RJMCMC algorithm will be a Gibbs-style algorithm. In addition to the full posterior distribution in (9), Gibbs-style algorithms require the conditional posterior distribution for the coefficient matrix and intercept matrix. Therefore, the posterior distribution in (9) together with the following posterior distribution will be employed in the MCMC estimation procedure:

$$p\left(\mathbf{A}, C|\mathbf{X}, \mathcal{N}_j\right) \propto (2\pi)^{-\frac{np}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left[\text{vec}\left(\mathbf{X}^T\right)\right]^T \left[\mathbf{H}^{-1}\right]\left[\text{vec}\left(\mathbf{X}^T\right)\right]\right]. \quad (10)$$

We can now design the required posterior simulator.

3 Posterior simulator

Outlined below is the RJMCMC scheme for estimating the parameters of the BEKK model. The models to be included in the search are a constant conditional covariance model, a diagonal BEKK model and a full BEKK model.

3.1 M-GARCH: main algorithm

Algorithm 1 M-GARCH - Main Algorithm

Require: $\left(\mathcal{N}_j^{(0)}, \mathbf{A}^{(0)}, C^{(0)}\right)$ either randomly or deterministically.

  for $i = 1 \ldots K$ **do**

    RJ$_\mathcal{N}$ Step - Simulate the GARCH model index, the intercept matrix and coefficient parameter matrix using $\left(\mathcal{N}_j^{(i)}, \mathbf{A}^{(i)}, C^{(i)}\right) \sim \mathrm{p}\left(\mathcal{N}_j, \mathbf{A}, C | \mathbf{X}\right)$ shown in (9)

    if RJ$_\mathcal{N}$ Step is rejected **then**

      DR$_\mathbf{A}$ Step - Simulate the intercept and coefficient parameter matrices from the full conditional posterior distribution: $\left(\mathbf{A}^{(i)}, C^{(i)}\right) \sim \mathrm{p}\left(\mathbf{A}, C | \mathbf{X}, \mathcal{N}_j^{(i)}\right)$ shown in (10).

    end if

  end for

Next, we expand the steps from the main algorithm that employ the RJMCMC and DRMH algorithms.

3.2 Reversible jump: RJ$_\mathcal{N}$ step

Step 2(a) in the main algorithm will be a Reversible Jump step as the move to another model index changes the dimension of the parameter space. The acceptance probability for a Reversible Jump step where the complete parameter matrix is proposed and there is no transformation from one space to the other will be

$$r_{(\mathcal{N}_j, j \to j*)} = \min\left(1, \frac{\mathrm{p}\left(\mathcal{N}_{j*}, \mathbf{A}^*, C^* | \mathbf{X}\right) \mathrm{q}\left(\mathbf{A}, C | \mathbf{A}^*, C^*, \mathcal{N}_j, \mathcal{N}_{j*}\right)}{\mathrm{p}\left(\mathcal{N}_j, \mathbf{A}, C | \mathbf{X}\right) \mathrm{q}\left(\mathbf{A}^*, C^* | \mathbf{A}, C, \mathcal{N}_j, \mathcal{N}_{j*}\right)}\right). \tag{11}$$

The proposals for the candidate intercept and coefficient matrices will be obtained separately. If we assume that $\mathbf{A}$ and C are independent, the proposal distributions will have the following structures:

$$\mathrm{q}\left(\mathbf{A}^*, C^* | \mathbf{A}, C, \mathcal{N}_j, \mathcal{N}_{j*}\right) = \mathrm{q}\left(\mathbf{A}^* | \mathbf{A}, \mathcal{N}_j, \mathcal{N}_{j*}\right) \mathrm{q}\left(C^* | C, \mathcal{N}_j, \mathcal{N}_{j*}\right).$$

As the intercept matrix must be positive definite, the proposal distribution for C will be a Wishart distribution. There are two options for the coefficient matrices. The first is to simply use separate Matrix Normal distributions for each A_i and B_j. While this is perfectly acceptable, it results in the multiplication of $l + m$ densities when models of orders $l, m > 1$ are allowed. This will add complexity to the coding of the algorithm.

Therefore, for the full BEKK coefficient matrix, the complete matrix $\mathbf{A}^*$ will be proposed from a Matrix Normal distribution. The dimension of this Matrix Normal distribution is $p \times p\,(l + m)$. The derivation of the mathematical form of this proposal distribution for the coefficient matrix is provided in Livingston (2017). The structure of this derivation indicates that proposing the full coefficient parameter matrix is

equivalent to proposing the parameter matrices separately, assuming the among row covariance is equal for each parameter matrix, that is $U_1 = U_2 = \cdots = U_r$.

For the proposal of the elements of the DBEKK model coefficient parameter matrix, the diagonal elements are simply combined into a single vector and a Multivariate Normal distribution is used.

It is essential to the success of the algorithm to have reasonable estimates of the location parameters of the proposal distributions. A simple and effective option is to perform a short pilot run to find location parameters that will allow the full algorithm to make model moves under the Reversible Jump scheme.

In general, the proposal distributions for the M-GARCH Reversible Jump step will be

$$\text{q}\left(\mathbf{A}^*|\mathbf{A}, \mathcal{N}_j, \mathcal{N}_{j*}\right) = 1 \qquad \text{(Constant)}$$

$$\text{q}\left(\mathbf{A}^*|\mathbf{A}, \mathcal{N}_j, \mathcal{N}_{j*}\right) \sim \mathcal{N}_{l+m}\left(\boldsymbol{m}_{\mathcal{N}_{j*}}, V_{\mathcal{N}_{j*}}\right) \qquad \text{(DBEKK)}$$

$$\text{q}\left(\mathbf{A}^*|\mathbf{A}, \mathcal{N}_j, \mathcal{N}_{j*}\right) \sim \mathcal{N}_{p, p(l+m)}\left(\mathbf{M}_{\mathcal{N}_{j*}}, U_{\mathcal{N}_{j*}}, \mathbf{V}_{\mathcal{N}_{j*}}\right) \qquad \text{(BEKK)} \qquad (12)$$

$$\text{q}\left(C^*|C, \mathcal{N}_j, \mathcal{N}_{j*}\right) \sim \mathcal{W}_p\left(\Delta_{C_{\mathcal{N}_{j*}}}^{-1} C_{\mathcal{N}_{j*}}, \Delta_{C_{\mathcal{N}_{j*}}}\right).$$

The Wishart distribution in (12) has a location parameter of $C_{\mathcal{N}_{j*}}$. This location parameter, along with the location parameters for the proposal distribution of the conditional covariance coefficient parameter matrix, $\mathbf{M}_{\mathcal{N}_{j*}}$ or $\boldsymbol{m}_{\mathcal{N}_{j*}}$, is determined during the pilot run. We have given the parameters the subscript $\mathcal{N}_{j*}$ to remind the reader that these parameters have the potential to be different for each model index. The variance parameters $U_{\mathcal{N}_{j*}}, \mathbf{V}_{\mathcal{N}_{j*}}$, and $\Delta_{C_{\mathcal{N}_{j*}}}$ can roughly be determined during the pilot run, but may require some adjustment to ensure an adequate acceptance rate for the algorithm.

The proposal distributions shown in (12) for the conditional covariance coefficient parameter matrix pertain to the general case in which the RJMCMC scheme allows not only jumps from different model types, but also jumps to differing model orders within a particular model. For practical reasons, the estimation algorithm implemented will only consider the models shown in Table 1. In particular, the model orders l and m will be fixed.

When a move to a candidate model is proposed, the proposal distributions for $\mathbf{A}$ for the candidate and current models will differ. This means that at each Reversible Jump, the acceptance probability can take a total of six different mathematical forms, depending on the indices of the current and candidate models. Each of these acceptance probabilities is will take the form as in (11).

The algorithm for the RJ$_{\mathcal{N}}$ Step is shown below.

Algorithm 2 RJ$_\mathcal{N}$ Algorithm

Ensure: $\left(\mathcal{N}_j^{(i-1)}, \mathbf{A}^{(i-1)}, C^{(i-1)}\right)$.

  for iteration i **do**

    • Propose a candidate model index $\mathcal{N}_{j*}$, from j$\left(\mathcal{N}_{j*}|\mathcal{N}_j\right)$ using the discretised Laplacian shown in Livingston and Nur (2017). Propose a candidate intercept and coefficient matrices using the proposal distributions in (12).

    • Determine the appropriate acceptance probability, based on $\mathcal{N}_j$ and $\mathcal{N}_{j*}$, and calculate the acceptance probability, $r_{\left(\mathcal{N}_j, j \to j*\right)}$, using (11).

    • Simulate $u \sim \mathcal{U}(0, 1)$:

    if $r_{\left(\mathcal{N}_j, j \to j*\right)} > u$ **then**

      • Accept the candidate model index $\mathcal{N}_{j*}$ and set $\mathcal{N}_{j*}^{(i)} = \mathcal{N}_{j*}$.

      • Accept the intercept and coefficient parameter matrices. Set $C^{(i)} = C^*$ and $\mathbf{A}^{(i)} = \mathbf{A}^*$.

    else

      • Reject the candidate model index $\mathcal{N}_{j*}$ and proposed covariance parameter matrices. Set $\mathcal{N}_j^{(i)} = \mathcal{N}_j^{(i-1)}$, $C^{(i)} = C^{(i-1)}$ and $\mathbf{A}^{(i)} = \mathbf{A}^{(i-1)}$.

      • Using the distribution in (10), simulate the intercept and coefficient parameter matrices,

$$(\mathbf{A}, C)^{(i)} \sim \mathrm{p}\left(\mathbf{A}, C|\mathbf{X}, \mathcal{N}_j^{(i)}\right),$$

  using the DRMH algorithm described below (DR$_\mathbf{A}$ Step).

    end if

  end for

3.3 Delayed rejection metropolis-hastings: DR$_\mathbf{A}$ Step

When the proposed jump to a new model index is rejected, the intercept and coefficient matrices need to be simulated from their joint conditional posterior distribution. If this did not occur, the intercept and coefficient matrices would only be updated upon model moves. When the data presents significant evidence for one model over the others included in the search, it is possible that model moves will not often take place. If the intercept and coefficient matrices are not separately updated, the algorithm will exhibit poor mixing and slow convergence.

Slow convergence of the main algorithm is a significant problem faced in this research. Within the univariate GARCH setting, it was possible to use maximum likelihood estimates as starting values for the algorithm. However, in the multivariate setting, finding the maximum likelihood estimates is not a simple task in itself and thus it is a reasonably difficult problem to identify suitable starting values quickly. Given that our starting values are likely to be relatively poor, we need to find techniques that allow our algorithms to move quickly towards the area of convergence. One such technique that we have found successful is the DRMH algorithm. The DRMH algorithm allows the use of relatively large variances for the proposal distributions so that initially the algorithm moves quickly towards the area of convergence, but also allows small movements if a poor proposal is made in the wrong direction, or even past the area of convergence.

Here we shall implement a two stage DRMH algorithm. This involves the prescription of proposal distributions for both stages of the algorithm. The proposal

distributions used in the first stage will be similar to those used for the Reversible Jump step above, but they will be centred on the most recent value in the MCMC chain. The proposal distributions employed are

$$
\begin{aligned}
q_{1\mathbf{A}}\left(\mathbf{A}^*|\mathbf{A}\right) &\sim \mathcal{N}_{p,l+m}\left(\mathbf{A}, U_{\mathcal{N}_j}, \mathbf{V}_{\mathcal{N}_j}\right) \\
q_{1C}\left(C^*|C\right) &\sim \mathcal{W}_p\left(\Delta_{C_{\mathcal{N}_j}}^{-1} C, \Delta_{C_{\mathcal{N}_j}}\right).
\end{aligned}
\tag{13}
$$

For the second stage of the proposal, our distributions will use both the information contained within the most recently simulated parameter and that contained within the candidate parameter that was rejected in the first stage. This gives rise to the following proposal distributions:

$$
\begin{aligned}
q_{2\mathbf{A}}\left(\mathbf{A}^{'}|\mathbf{A}^*, \mathbf{A}\right) &\sim \mathcal{N}_{p,l+m}\left(\frac{1}{2}\left(\mathbf{A}^* + \mathbf{A}\right), U_{\mathcal{N}_j}, \Delta_{2\mathbf{A}}\mathbf{V}_{\mathcal{N}_j}\right) \\
q_{2C}\left(C^{'}|C^*, C\right) &\sim \mathcal{W}_p\left(\left[2\Delta_{C_{\mathcal{N}_j}}\Delta_{2C}\right]^{-1}\left(C^* + C\right), \Delta_{C_{\mathcal{N}_j}}\Delta_{2C}\right),
\end{aligned}
\tag{14}
$$

where Δ_{2C} and $\Delta_{2\mathbf{A}}$ are parameters included to reduce the variance of the proposal in the second stage. Setting these values in the range of 0.50 to 0.75 yielded successful results during testing of the algorithm.

The target distribution for this step is the joint conditional posterior distribution shown in (10). That is,

$$
q\left(\mathbf{A}, C|\mathbf{X}, \mathcal{N}_j\right) \propto (2\pi)^{-\frac{np}{2}}|\mathbf{H}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left[\operatorname{vec}\left(\mathbf{X}^T\right)\right]^T\left[\mathbf{H}^{-1}\right]\left[\operatorname{vec}\left(\mathbf{X}^T\right)\right]\right].
$$

Since there are three different situations in which this algorithm could be implemented, namely for the constant conditional covariance model, the DBEKK model and the BEKK model, we only present the most complicated form of the acceptance probability. The difference between the possible forms ultimately lies in the ratio of the proposal distributions for the conditional covariance coefficient parameter matrices.

The acceptance probability for the first stage of the DRMH algorithm is relatively simple. It is simply a Metropolis-Hastings acceptance probability:

$$
r_{(\mathbf{A},1)}\left(\mathbf{A}^*, C^*, \mathbf{A}, C\right) = \min\left(1, \frac{p\left(\mathbf{A}^*, C^*|\mathbf{X}, \mathcal{N}_j\right)}{p\left(\mathbf{A}, C|\mathbf{X}, \mathcal{N}_j\right)}\frac{q_{1\mathbf{A}}\left(\mathbf{A}|\mathbf{A}^*\right)q_{1C}\left(C|C^*\right)}{q_{1\mathbf{A}}\left(\mathbf{A}^*|\mathbf{A}\right)q_{1C}\left(C^*|C\right)}\right). \tag{15}
$$

In cases where the initial proposed parameters are rejected, the $DR_{\mathbf{A}}$ Algorithm enters the second stage. This stage requires a slightly more complicated acceptance probability to ensure convergence to the correct stationary distribution. The required acceptance probability has the following form:

$$r_{(A,2)} \left(A^{'}, C^{'}, A^{*}, C^{*}, A, C \right) =$$

$$\min \left(1, \frac{p \left(A^{'}, C^{'} | X, \mathcal{N}_j \right)}{p \left(A, C | X, \mathcal{N}_j \right)} \frac{q_{1A} \left(A^{*} | A^{'} \right) q_{1C} \left(C^{*} | C^{'} \right)}{q_{1A} \left(A^{*} | A \right) q_{1C} \left(C^{*} | C \right)} \times \right.$$

$$\left. \frac{q_{2A} \left(A | A^{'}, A^{*} \right) q_{2C} \left(C | C^{'}, C^{*} \right)}{q_{2A} \left(A^{'} | A, A^{*} \right) q_{2C} \left(C^{'} | C, C^{*} \right)} \frac{\left(1 - r_1 \left(A^{*}, C^{*}, A^{'}, C^{'} \right) \right)}{\left(1 - r_1 \left(A^{*}, C^{*}, A, C \right) \right)} \right) . \tag{16}$$

The layout of the DR$_A$ Algorithm is outlined below.

Algorithm 3 DR$_A$ Algorithm

Ensure: $\left(\mathcal{N}_j^{(i)}, A^{(i-1)}, C^{(i-1)} \right)$.

  for iteration i **do**

  • Propose new intercept and coefficient parameter matrices C^{*} and A^{*} from the proposal densities, $q_{1C} \left(C^{*} | C \right)$ and $q_{1A} \left(A^{*} | A \right)$, respectively which are given in (13).

  • Calculate the acceptance probability $r_{(A,1)} \left(A^{*}, C^{*}, A, C \right)$ using the formula in (15).

  • Simulate $u_1 \sim \mathcal{U} (0, 1)$:

  if $r_{(A,2)} \left(A^{'}, C^{'}, A^{*}, C^{*}, A, C \right) > u_2$ **then**

    • Accept the new proposed intercept and coefficient parameter matrices. Set $C^{(i)} = C^{'}$ and $A^{(i)} = A^{'}$.

  else

    • Generate new proposed intercept and coefficient parameter matrices, $C^{'}$ and $A^{'}$ from $q_{2C} \left(C^{'} | C^{*}, C \right)$ and $q_{2A} \left(A^{'} | A^{*}, A \right)$, respectively, using the proposal distributions in (14).

    • Calculate the acceptance probability, $r_{(A,2)} \left(A^{'}, C^{'}, A^{*}, C^{*}, A, C \right)$, using the formula in (16).

    • Simulate $u_2 \sim \mathcal{U} (0, 1)$:

    if $r_{(A,2)} \left(A^{'}, C^{'}, A^{*}, C^{*}, A, C \right) > u_2$ **then**

      • Accept the new proposed intercept and coefficient parameter matrices. Set $C^{(i)} = C^{'}$ and $A^{(i)} = A^{'}$.

    else

      • Reject the new proposed intercept and coefficient parameter matrices. Set $C^{(i)} = C^{(i-1)}$ and $A^{(i)} = A^{(i-1)}$.

    end if

  end if

  end for

It is relatively straight forward to extend the DRMH algorithm above to include more stages. With the algorithms elucidated, the following section outlines simulations studies confirming their effectiveness.

4 Empirical analysis

4.1 Simulation study

The three models included in the search are the fixed conditional covariance model, the DBEKK and the full BEKK formulations of the conditional covariance equation in an M-GARCH model. These are indexed using $\mathcal{N}_j$, as outlined in Table 1.

Two simulation studies are presented with the first simulation study denoted as MG-I; where $N = 1,500$. The following true DBEKK model was used:

$$x_t = \varepsilon_t, \quad \varepsilon_t = H_t^{\frac{1}{2}} \eta_t$$

$$H_t = \begin{bmatrix} 0.20 & -0.05 & 0.25 \\ -0.05 & 0.30 & 0.00 \\ 0.25 & 0.00 & 0.60 \end{bmatrix} + \begin{bmatrix} 0.70 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.75 \end{bmatrix} \varepsilon_{t-1} \varepsilon_{t-1}^T \begin{bmatrix} 0.70 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.75 \end{bmatrix}^T +$$

$$\begin{bmatrix} 0.55 & 0.00 & 0.00 \\ 0.00 & 0.65 & 0.00 \\ 0.00 & 0.00 & 0.45 \end{bmatrix} H_{t-1} \begin{bmatrix} 0.55 & 0.00 & 0.00 \\ 0.00 & 0.65 & 0.00 \\ 0.00 & 0.00 & 0.45 \end{bmatrix}^T .$$

An example of simulated data from MG-I is shown in Fig. 1. The pilot runs for each of the three model indices were executed for differing numbers of iterations. The simpler models, being the constant conditional covariance and the DBEKK model, generally converged to the area of high posterior probability fairly quickly, when compared with the full BEKK model. With this in mind, the lengths of the pilot runs were set to 1,000, 4,000 and 12,000 iterations for the constant, DBEKK and BEKK models, respectively.

The main algorithm that allows jumps between models was run for 50,000 iterations, of which the first 10,000 iterations were discarded as a burn in. The results shown in Table 2 indicate that the true model was correctly identified 977 times out of the 1,000 replications. In the other 23 runs, a full BEKK model was identified. The means of the parameter estimates from the runs that incorrectly identified the BEKK model were similar to those for the true DBEKK model.

The results in Table 2 show that the posterior mean estimates were very close to the true values. In addition, the coverage probability was close to the notional interval value of 95%. A set of posterior distributions from one of the estimation runs is shown in Fig. 2. The posterior distribution for the model order is omitted as the posterior probability was 100% at the true value for this application of the estimation scheme.

For the second study, MG-II, the data length was set to $N = 1500$. The example of simulated data shown in Fig. 3 was created from the following true BEKK model:

$$x_t = \varepsilon_t, \quad \varepsilon_t = H_t^{\frac{1}{2}} \eta_t,$$

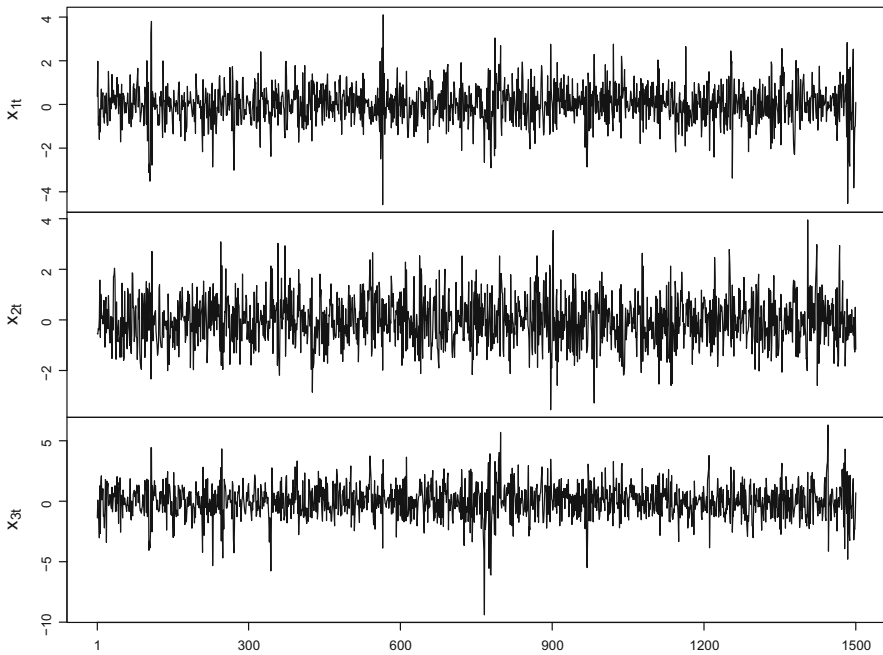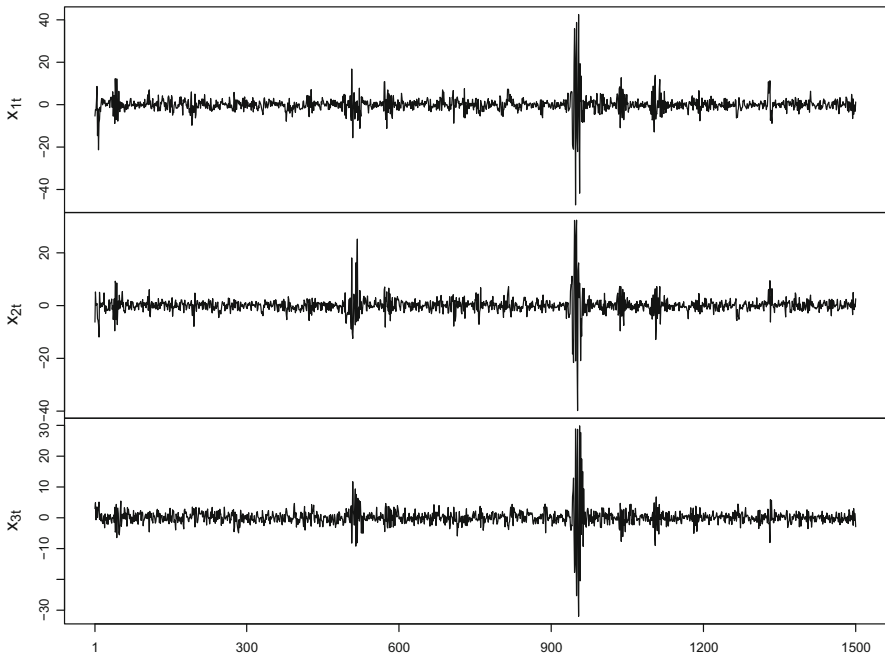with conditional covariance equation

Fig. 1 Time plot of simulated M-GARCH data x_t from MG-I

Table 2 Summary statistics for the simulated M-GARCH model of MG-I

Parameter	True value	Estimate	CrI	Coverage (%)
A [1, 1]	0.70	0.7025	(0.6555, 0.7512)	94.68
A [2, 2]	0.50	0.5028	(0.4482, 0.5590)	94.78
A [3, 3]	0.75	0.7528	(0.7040, 0.8032)	95.91
B [1, 1]	0.55	0.5422	(0.4864, 0.5951)	95.29
B [2, 2]	0.65	0.6303	(0.5311, 0.7182)	94.17
B [3, 3]	0.45	0.4415	(0.3828, 0.4987)	94.68
C [1, 1]	0.20	0.2053	(0.1722, 0.2412)	94.17
C [2, 1]	-0.05	-0.0520	$(-0.0732, -0.0322)$	95.09
C [3, 1]	0.25	0.2570	(0.2165, 0.3009)	94.06
C [2, 2]	0.30	0.3166	(0.2334, 0.4067)	93.96
C [2, 3]	0.00	0.0005	$(-0.0326, 0.0337)$	94.88
C [3, 3]	0.60	0.6137	(0.5292, 0.7042)	94.06
$\mathcal{N}_j$	2	2	–	97.70
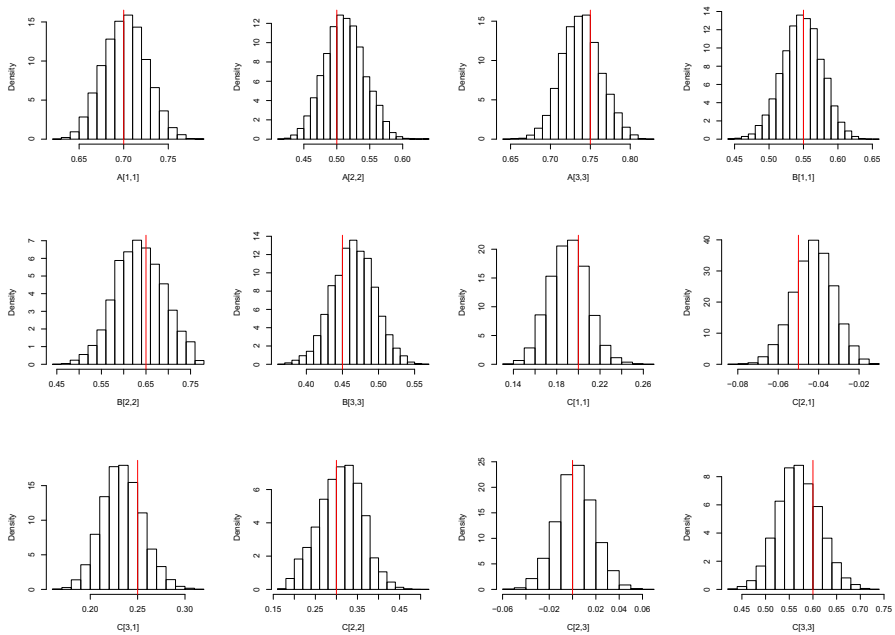
Fig. 2 An example set of parameter posterior distributions from MG-I. The red lines indicate the positions of the true values for each parameter

Fig. 3 Time plot of simulated M-GARCH data x_t from MG-II

Table 3 Summary statistics for the simulated M-GARCH model of MG-II

Parameter	True value	Estimate	CrI	Coverage (%)
$A[1, 1]$	0.68	0.6820	(0.6306 0.7350)	95.0
$A[2, 1]$	0.35	0.3541	(0.3169 0.3924)	94.5
$A[3, 1]$	-0.25	-0.2510	$(-0.2857 -0.2170)$	92.6
$A[1, 2]$	0.18	0.1843	(0.1257 0.2433)	93.4
$A[2, 2]$	0.50	0.4964	(0.4463 0.5480)	95.1
$A[3, 2]$	0.12	0.1192	(0.0750 0.1639)	93.3
$A[1, 3]$	0.40	0.4008	(0.3415 0.4611)	92.2
$A[2, 3]$	0.00	0.0019	$(-0.0454\ 0.0492)$	93.8
$A[3, 3]$	0.35	0.3474	(0.2993 0.3963)	94.0
$B[1, 1]$	0.55	0.5431	(0.4971 0.5872)	91.0
$B[2, 1]$	-0.15	-0.1511	$(-0.1901 -0.1117)$	92.3
$B[3, 1]$	0.25	0.2445	(0.2061 0.2836)	90.6
$B[1, 2]$	0.20	0.1968	(0.1175 0.2771)	91.0
$B[2, 2]$	0.60	0.6000	(0.5413 0.6562)	91.0
$B[3, 2]$	-0.40	-0.3978	$(-0.4614\ -0.3347)$	88.9
$B[1, 3]$	-0.20	-0.1949	$(-0.2641\ -0.1280)$	93.3
$B[2, 3]$	0.30	0.2909	(0.2287 0.3555)	91.2
$B[3, 3]$	0.65	0.6394	(0.5749 0.7025)	89.7
$C[1, 1]$	0.56	0.5704	(0.4079 0.7468)	94.5
$C[2, 1]$	0.19	0.1906	(0.0565 0.3320)	91.0
$C[3, 1]$	0.08	0.0933	$(-0.0353\ 0.2212)$	91.2
$C[2, 2]$	0.47	0.4690	(0.2997 0.6469)	91.1
$C[2, 3]$	0.21	0.2321	(0.0914 0.3618)	87.9
$C[3, 3]$	0.42	0.4202	(0.2660 0.5829)	89.5
$\mathcal{N}_j$	3	3	–	100.00

$$H_t = \begin{bmatrix} 0.56 & 0.19 & 0.08 \\ 0.19 & 0.47 & 0.21 \\ 0.08 & 0.21 & 0.42 \end{bmatrix} + \begin{bmatrix} 0.68 & 0.18 & 0.40 \\ 0.35 & 0.50 & 0.00 \\ -0.25 & 0.12 & 0.35 \end{bmatrix} \varepsilon_{t-1}\varepsilon_{t-1}^T \begin{bmatrix} 0.68 & 0.18 & 0.40 \\ 0.35 & 0.50 & 0.00 \\ -0.25 & 0.12 & 0.35 \end{bmatrix}^T +$$

$$\begin{bmatrix} 0.55 & 0.20 & -0.20 \\ -0.15 & 0.60 & 0.30 \\ 0.25 & -0.40 & 0.65 \end{bmatrix} H_{t-1} \begin{bmatrix} 0.55 & 0.20 & -0.20 \\ -0.15 & 0.60 & 0.30 \\ 0.25 & -0.40 & 0.65 \end{bmatrix}^T .$$

Comparing the plots in Figs. 1 and 3, it is evident that the full BEKK formulation of the conditional covariance equation allows for shocks to spill over to other series. Areas of high volatility are shared by each of the plots.

The pilot runs for each of the three types of models were the same as for MG-I. The main algorithm, however, was run for 150,000 iterations, of which the first 10,000 were discarded as a burn in. A typical set of posterior distributions from an individual run of the algorithm is shown in Fig. 4.
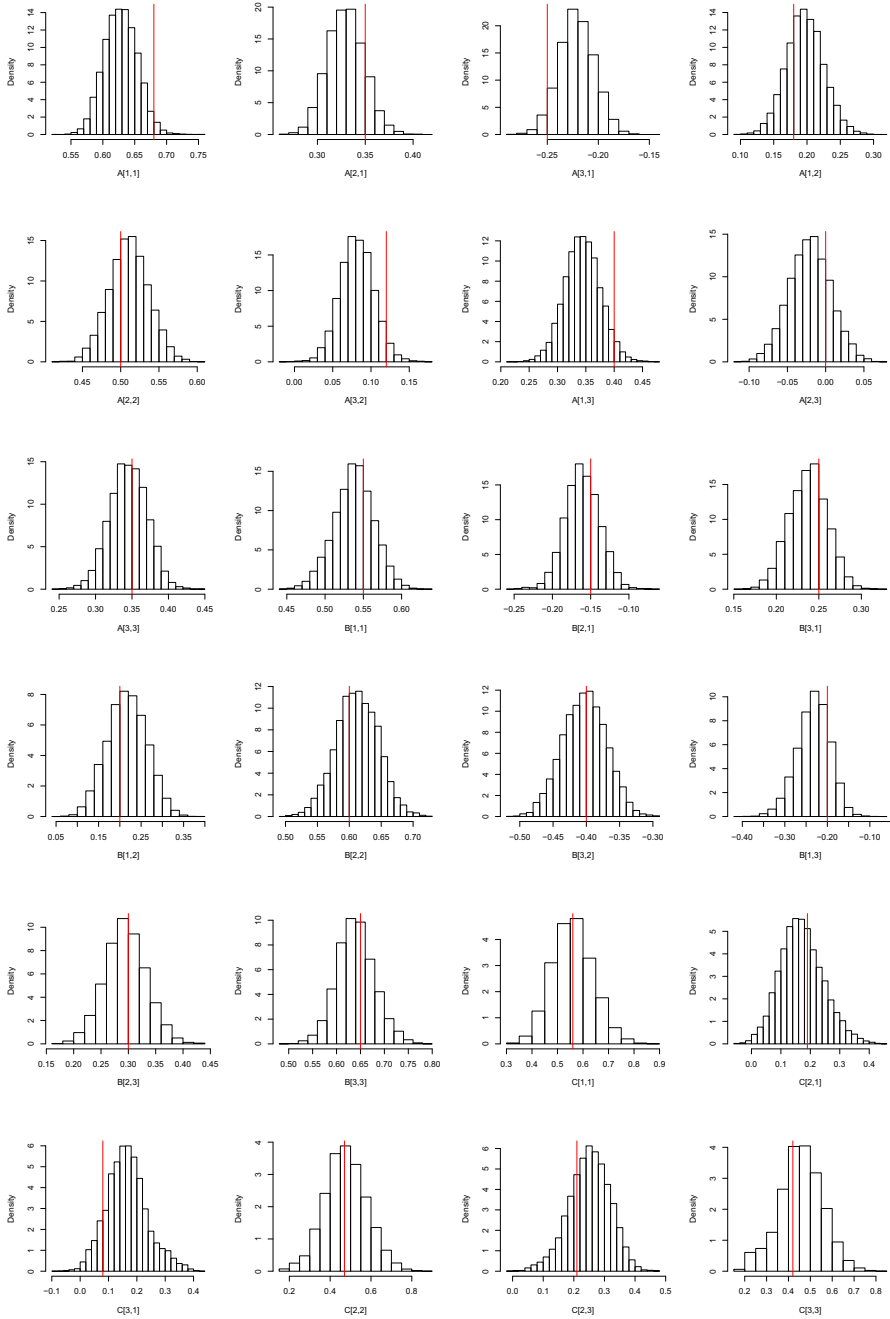
Fig. 4 An example set of parameter posterior distributions from MG-II. The red lines indicate the positions of the true values for each parameter

The results shown in Table 3 indicate that the means of the parameter point estimates were very close to their true values at two decimal places. The coverage probabilities indicate that the credible intervals underperformed: most coverage probabilities were less than 95%. There were a few parameters for which the coverage probability was less than 90%. However, the mean coverage probability over all parameters except for the model index was 92%, which is not substantially different from the notional value of 95%.

The application of the M-GARCH estimation scheme to real world data is discussed in the following section.

4.2 Real world data

An illustration of the M-GARCH estimation algorithm was implemented using trivariate data derived from the daily closing prices for BHP Group Limited (BHP); Rio Tinto Group (RIO); and, Fortescue Metals Group Limited (FMG), all traded on the Australian Stock Exchange. The data ranges from September 2013 to December 2021. The data was cleaned to only include closing prices on dates for which all series had data. We acknowledge there are more sophisticated methods for dealing with missing data.

The closing prices are plotted in Fig. 5. From the plots it appears that BHP and RIO are highly correlated throughout the period. The correlations of BHP and RIO to FMG prior to May 2020 do not seem particularly high, however after May 2020 the correlation appears greater. The high correlation between BHP and RIO is to be somewhat expected given the similarities in business activities and size of the companies. While FMG is also a mining company, it is considerably smaller and likely focuses on different commodities.

Rather than analysing the series directly, the observations were first transformed into their log returns as follows:

$$x_{1,t}* = \log (\text{BHP}_t) - \log (\text{BHP}_{t-1})$$
$$x_{2,t}* = \log (\text{RIO}_t) - \log (\text{RIO}_{t-1})$$
$$x_{3,t}* = \log (\text{FMG}_t) - \log (\text{FMG}_{t-1}) .$$

Next the $x_{i,t}*$ were all standardised to create $x_{i,t}$ by subtracting the mean and dividing by the standard deviation of the respective series. The resulting time series have $N = 2,047$ observations. A time plot of the data that was analysed is shown in Fig. 6. Comparing the conditional variance across the three series, all three exhibit increases in conditional variance between May 2015 and January 2017, then again before May 2020.

The sample variances of a rolling window of length ten were calculated for the series and are presented in Fig. 7. The plot confirms the periods of increased conditional variance and also indicates that for FMG, there is increased conditional variance prior to May 2015. It appears that a DBEKK or full BEKK multivariate GARCH model would be well suited for modelling the data, and therefore these models are used as candidates to model the conditional covariance.
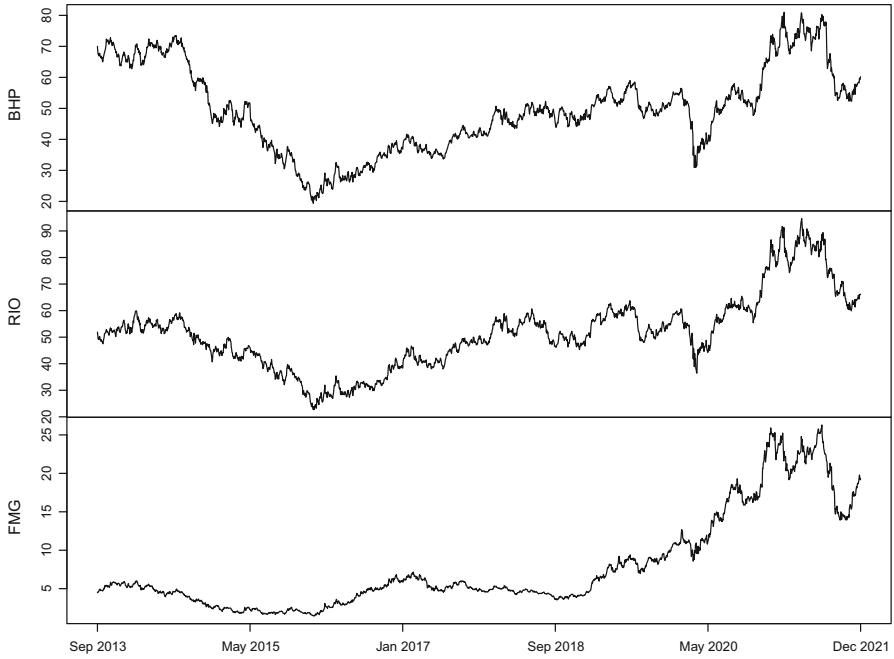
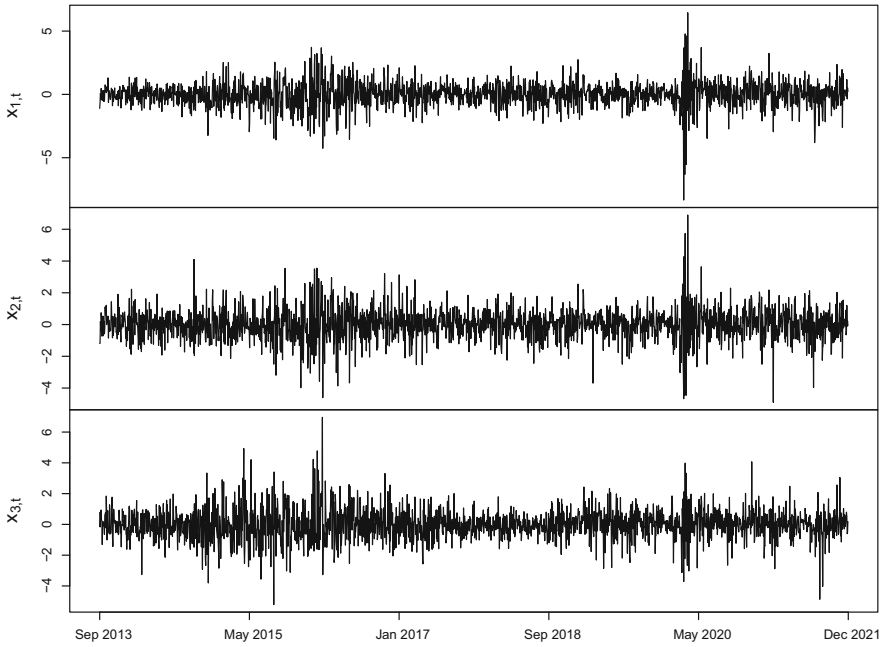**Fig. 5** Daily closing price BHP, RIO, and FMG share prices

Fig. 6 Time plots of the log returns $x_{1,t}$, $x_{2,t}$, and $x_{3,t}$, for BHP, RIO, and FMG respectively
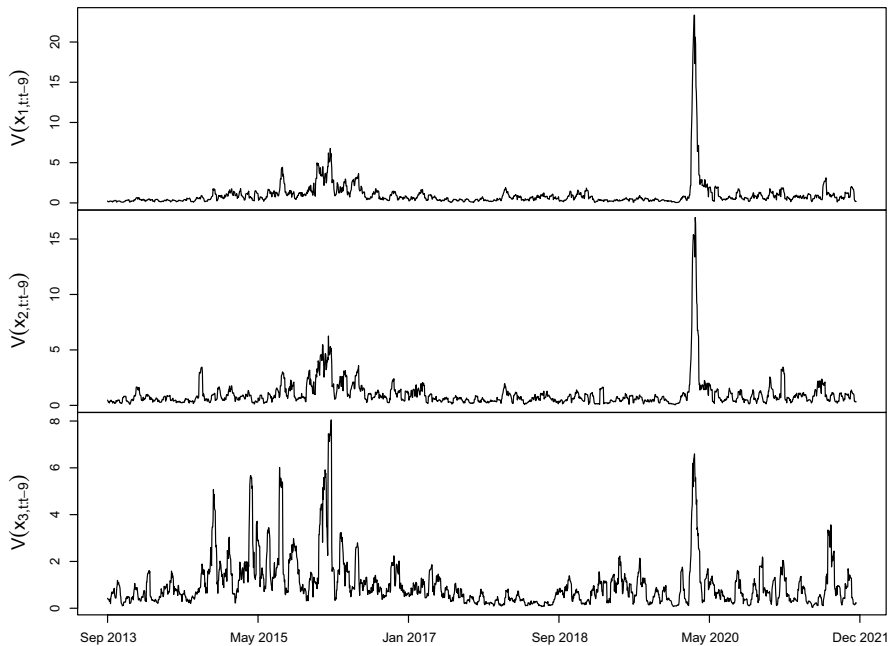
Fig. 7 Sample variances of a rolling window of length 10, for BHP, RIO, and FMG

The pilot runs of the algorithm were executed for 20,000 iterations for each of the three models under consideration. The first 10,000 iterations were used as burn in. The full algorithm was then executed for 200,000 iterations, again using the first 10,000 iterations as a burn in period.

The posterior evidence supporting the DBEKK formulation of the conditional covariance equation was extremely strong. The algorithm never departed from that model type after the first few iterations. Supporting this are the DIC values of 23,741 and 25,611 for the DBEKK and BEKK models respectively. The respective AIC values are 23,607 and 25,387, indicating that the DBEKK model is incomparably more likely. As a comparison, estimating the full BEKK model through maximum likelihood estimation using the `mgarchBEKK` package (Schmidbauer et al. 2016) in R resulted in a calculated AIC value of 27,320. Some effort was put into improving the initial values of the optimisation routine, however the improvements to the calculated AIC were not significant. Additionally, some sets of initial values resulted in warnings from the optim function related to the the calculation of the Hessian matrix which resulted in some standard errors being unavailable. Documentation for the package indicates that development ceased in 2016.

Plots of the posterior distributions for each of the parameters are shown in Fig. 8 and the estimation results are displayed in Table 4. Several of the posterior distributions are skewed and the distribution of C [3, 3] appears to be bi-modal. These observations cast doubt over the validity and usefulness of the DIC calculations.
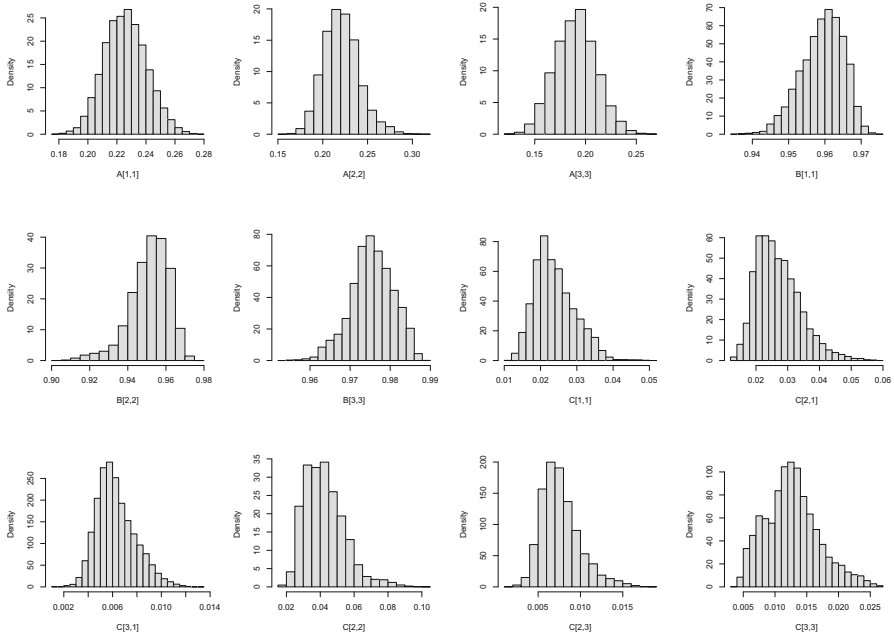
Fig. 8 Posterior distributions for the M-GARCH model fitted to the BHP, RIO, and FMG data

Table 4 Point estimates and credible intervals for the BHP, RIO, and FMG data

Parameter	Estimate	CrI	Parameter	Estimate	CrI
A [1, 1]	0.2260	(0.1989, 0.2553)	B [1, 1]	0.9593	(0.9468, 0.9687)
A [2, 2]	0.2208	(0.1852, 0.2656)	B [2, 2]	0.9514	(0.9248, 0.9671)
A [3, 3]	0.1903	(0.1519, 0.2310)	B [3, 3]	0.9755	(0.9640, 0.9850)
C [1, 1]	0.0239	(0.0149, 0.0358)	C [2, 1]	0.0270	(0.0164, 0.0436)
C [3, 1]	0.0063	(0.0037, 0.0098)	C [2, 2]	0.0425	(0.0252, 0.0728)
C [2, 3]	0.0076	(0.0042, 0.0134)	C [3, 3]	0.0127	(0.0055, 0.0224)

The fitted conditional covariance and correlations over the time period are shown in Fig. 9. The correlation plot between BHP and RIO shows that, over time, the correlation was quite high and varied between a minimum of 0.60 and a maximum of 0.95. The middle 95% ranged from 0.74 to 0.92, with a mean correlation of 0.84. The correlations of FMG to BHP and RIO are much less, however are almost always positive.

The mathematical form of the fitted model is
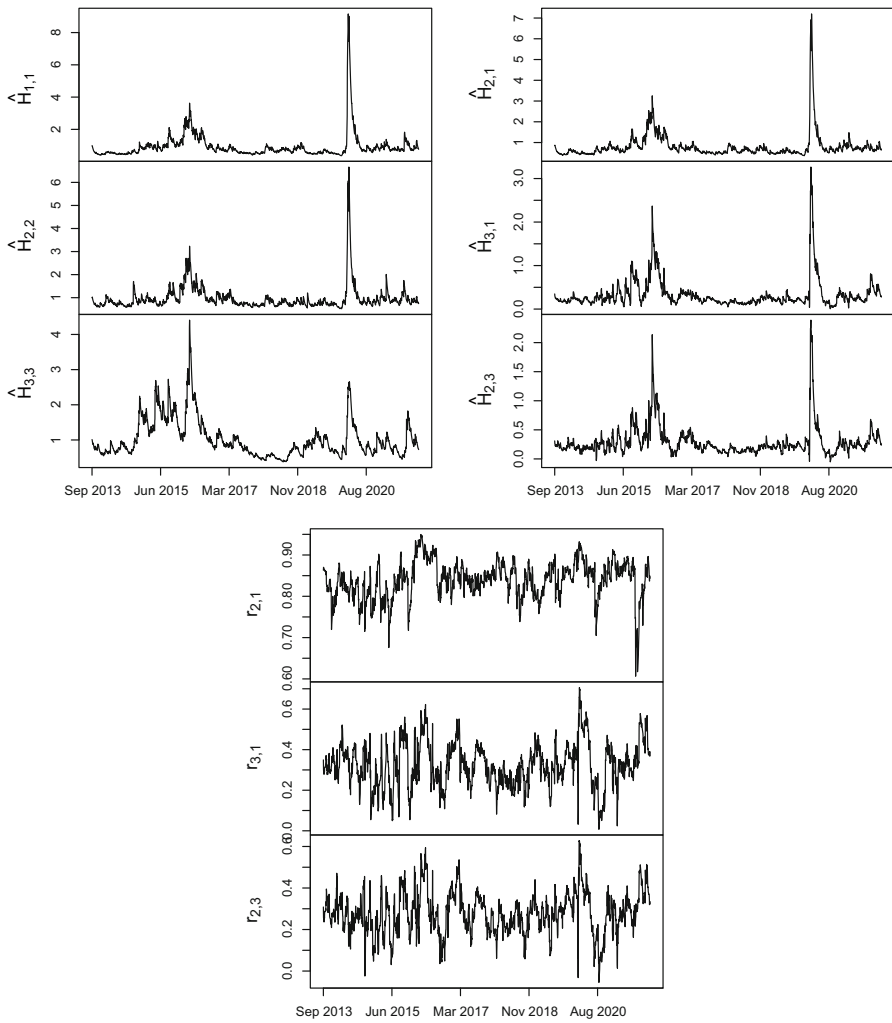
Fig. 9 Estimated conditional covariances (top) and correlations (bottom) over the time period, for the BHP, RIO, and FMG data

$$x_t = \varepsilon_t$$

$$\varepsilon_t = H_t^{\frac{1}{2}} \eta_t$$

$$H_t = \begin{bmatrix} 0.024 & 0.027 & 0.006 \\ 0.027 & 0.043 & 0.008 \\ 0.006 & 0.008 & 0.013 \end{bmatrix} + \begin{bmatrix} 0.226 & 0.000 & 0.00 \\ 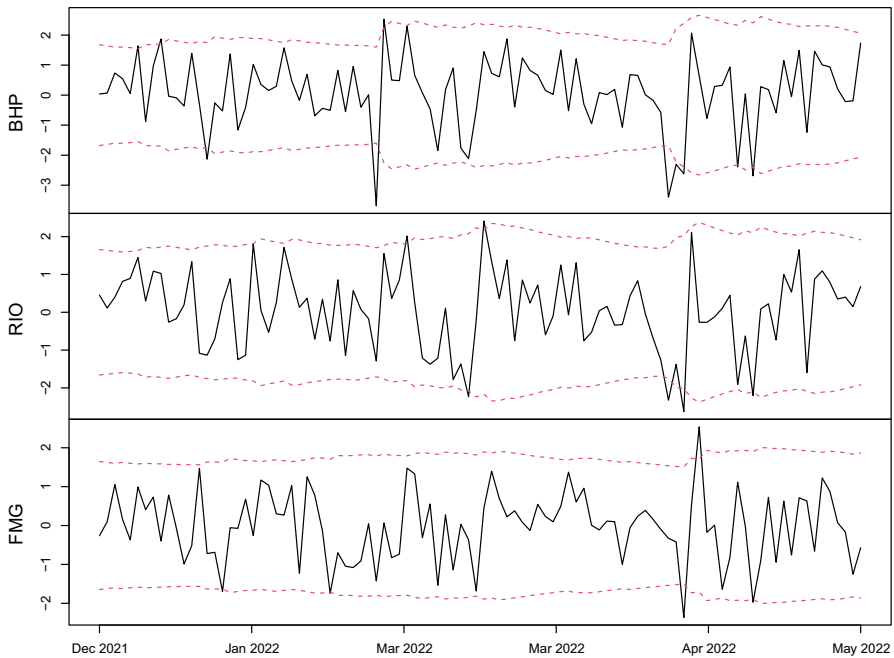0.000 & 0.221 & 0.00 \\ 0.000 & 0.000 & 0.19 \end{bmatrix} \varepsilon_{t-1} \varepsilon_{t-1}^T \begin{bmatrix} 0.226 & 0.000 & 0.00 \\ 0.000 & 0.221 & 0.00 \\ 0.000 & 0.000 & 0.19 \end{bmatrix}^T +$$

$$\begin{bmatrix} 0.959 & 0.000 & 0.000 \\ 0.000 & 0.951 & 0.000 \\ 0.000 & 0.000 & 0.975 \end{bmatrix} H_{t-1} \begin{bmatrix} 0.959 & 0.000 & 0.000 \\ 0.000 & 0.951 & 0.000 \\ 0.000 & 0.000 & 0.975 \end{bmatrix}^T.$$

Fig. 10 One step ahead out of sample forecasts for 100 days for BHP, RIO, and FMG data

While forecasting is not the focus of this research, it is useful to demonstrate model forecasting. Assessing forecast performance in volatility models is somewhat difficult in that what we are forecasting, conditional covariance or correlation, is unobserved. Therefore we do not have true values to make comparisons against. A common approach to testing forecast performance of M-GARCH models is to look at a Model Confidence Set (Hansen et al. 2003, 2011; Laurent et al. 2012; Caporin and McAleer 2014). This methodology is useful when one has a large selection of competing models. The method unfortunately requires assessment of all competing models and comparisons made across several metrics.

A much simpler approach is to compare forecast conditional covariances or correlations to *estimates* of their true values. In this case, estimates of their true values are obtained by looking at a rolling window of the data and calculating sample covariances and correlations. While calculations on a rolling window are suitable for determining indicative values and trends, the values obtained are highly sensitive to the size of the rolling window.

To assess forecast performance, one-step ahead forecasts are made out of sample without updating the model each day for 100 days. The proportion of observations that are within a 95% prediction interval are 90%, 94%, and 95% for BHP, RIO, and FMG respectively leading to an overall mean proportion of 93%. A plot of the forecast prediction intervals are shown in Fig. 10. For comparison, the forecasts from the fitted model using maximum likelihood estimation through the `mgarchBEKK` package in R resulted in 100%, 76%, and 74% of observations within the 95% prediction intervals

for BHP, RIO, and FMG respectively and an overall percentage of 83%. It appears as though the fitted model underestimates the conditional variance.

5 Concluding remarks

In this article, we have performed an analysis of M-GARCH-BEKK models from a Bayesian perspective. The resulting estimation scheme combines parameter estimation and model selection into a single task by including the model variants of BEKK models and the coefficient parameters as parameters to estimate.

We have obtained that: (i) the design of a posterior simulator using a combination of MCMC algorithms which includes Reversible Jump MCMC (RJMCMC) and Delayed Rejection Metropolis-Hastings (DRMH) algorithms respectively was computationally feasible to be implemented in R; (ii) the two-stage DRMH algorithm successfully resulted in faster convergence of the MCMC chains; (iii) the proposed method provides reliable estimates based on the performance of extensive simulation studies; (iv) uncertainties surrounding all parameters including the model variants are incorporated into the estimation procedure and finally (v) model selection has been embedded into the proposed method using a Bayesian approach.

It was found to be critically important to obtain good proposal distribution location parameters for the covariance coefficient parameter matrix for the M-GARCH model. This was achieved through pilot runs, which if not long enough, suitable location parameters were not always obtained. This meant that the algorithm was sometimes unable to jump to the most suitable model for the conditional covariance. This was much more important for real world data.

Due to the irregular shapes of the posterior distributions for the M-GARCH models, there was a risk that some MCMC chains would become trapped within local maxima. In order to reduce the chance of this occurring, the proposal variances needed to be set as large as possible, while still trying to maintain reasonable acceptance rates. The DRMH algorithm helped to alleviate this problem. A large number of iterations of the final stage of the algorithm was required in order to ensure convergence of the MCMC chains. Another reason for needing to run the algorithms for many iterations relates to the fact that the MCMC chains for the GARCH coefficient parameter matrices are highly correlated.

Consequently, this meant the run time of the algorithm was quite long, especially compared to the time taken to obtain maximum likelihood estimates from the `mgarchBEKK` package in R. The comparison here is not a fair one though as our procedure has been implemented in the statistical software package R without taking measures to optimise the speed of the code. The `mgarchBEKK` package employs the optim function in R which is coded in C, a significantly faster language.

For the purpose of comparing the speed of model selection and estimation from a Bayesian perspective, keeping programming language constant, assume that one wishes to extend the pool of candidate models to include other formulations of the covariance equation or include other types of models such as DCC and CCC. A typical approach to selecting a model such as DIC or Bayes Factors would require estimation of all candidate models followed by a comparison of the models using a chosen metric.

Using the RJMCMC algorithm requires a single estimation run to select the model and estimate the parameters. If the time taken to complete the pilot runs is less than the total time to estimate all of the candidate models, the algorithm presented here will be faster. This is likely as the pilot runs are relatively short as they are only needed to obtain location parameters.

Overall, the aims of the research were achieved, but some limitations and possible areas for future extensions were identified. To speed up the computation time of the algorithm, a faster programming language such as C or C++ should be used to improve the run time. This forms part of the first authors current research.

While some unsophisticated adaptive procedures were implemented when coding the algorithms, there are other, more effective adaptive techniques that would allow automatic tuning of the acceptance rates within the algorithms. This would eliminate the need for the user to control the acceptance rates by specifying the tuning parameters of the algorithms.

Another extension would be to include more classes of models in the searches. There are a plethora of time series models that could easily be incorporated into the model indices. For example, asymmetric covariance equations could be added to the index of possible models for the conditional covariance. A limitation of this research and an area available for extension is that the BEKK model orders have been fixed such that $l = m = Q = 1$. While this is common, there is very little literature exploring the advantages of not employing this restriction.

Furthermore, the error distributions used throughout this paper were the multivariate Normal distributions. Changing these to multivariate t-distributions for which the degrees of freedom are a parameter to be estimated would generalise the process even further. Some financial time series would be better suited to models that incorporate a fatter tailed error distribution.

References

Bauwens L, Hafner CM, Laurent S (2012) Handbook of volatility models and their applications, vol 3. Wiley, New York

Bauwens L, Laurent S, Rombouts JVK (2006) Multivariate Garch models: a survey. J Appl Econom 21(1):79–109

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econom 31(3):307–327

Bollerslev T, Engle RF, Wooldridge JM (1988) A capital asset pricing model with time-varying covariances. J Polit Econ 96(1):116–131

Caporin M, McAleer M (2014) Robust ranking of multivariate Garch models by problem dimension. Comput Stat Data Anal 76:172–185 (**CFEnetwork: The Annals of Computational and Financial Econometrics**)

Engle RF, Kroner KF (1995) Multivariate simultaneous generalized arch. Econom Theory 11:122–150

Francq C, Zakoïan J-M (2010) Multivariate GARCH processes. Wiley, New York, pp 273–310

Glosten LR, Jagannathan R, Runkle DE (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. J Financ 48(5):1779–1801

Green PJ (1995) Reversible jump Markov chain monte Carlo computation and Bayesian model determination. Biometrika 82(4):711–732

Hansen PR, Lunde A, Nason JM (2003) Choosing the best volatility models: the model confidence set approach*. Oxf Bull Econ Stat 65(s1):839–861

Hansen PR, Lunde A, Nason JM (2011) The model confidence set. Econometrica 79(2):453–497

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1):97–109

Jensen MJ, Maheu JM (2013) Bayesian semiparametric multivariate GARCH modeling. J Econom 176(1):3–17

Kearney C, Patton AJ (2000) Multivariate Garch modeling of exchange rate volatility transmission in the European monetary system. Financ Rev 35(1):29–48

Laurent S, Rombouts JVK, Violante F (2012) On the forecasting accuracy of multivariate Garch models. J Appl Econom 27(6):934–955

Li H (2007) International linkages of the Chinese stock exchanges: a multivariate Garch analysis. Appl Financ Econ 17(4):285–297

Livingston Jr G (2017) A Bayesian analysis of a regime switching volatility model. University of Newcastle Australia Research Higher Degree Thesis

Livingston G Jr, Nur D (2017) Bayesian inference for smooth transition autoregressive (STAR) model: a prior sensitivity analysis. Commun Stat 46(7):5440–5461

Livingston Jr G, Nur D (2018)Bayesian inference of smooth transition autoregressive (star)(k)–garch(l, m) models. *Stat Pap*

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21(6):1087–1092

Nelson DB (1991) Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59(2):347–370

R Development Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. 3-900051-07-0

Saleem K (2009) International linkage of the Russian market and the Russian financial crisis: a multivariate GARCH analysis. Res Int Bus Financ 23(3):243–256

Schmidbauer H, Roesch A, Tunalioglu VS (2016) mgarchBEKK: simulating, estimating and diagnosing MGARCH (BEKK and mGJR) processes. R package version 0.0.2

Silvennoinen A, Teräsvirta T (2009) Multivariate GARCH models. Springer, Berlin Heidelberg, pp, 201–229

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc B 64(4):583–639

Tierney L, Mira A (1999) Some adaptive Monte Carlo methods for Bayesian inference. Stat Med 18(17–18):2507–2515

Tsay RS (2013) Multivariate time series analysis: with R and financial applications. Wiley, New York

Vrontos ID, Dellaportas P, Politis DN (2000) Full Bayesian inference for Garch and Egarch models. J Bus Econ Stat 18(2):187–198

Vrontos ID, Dellaportas P, Politis DN (2003) Inference for some multivariate arch and Garch models. J Forecast 22(6–7):427–446